

請求項 1: X?  
9: X?

5-00059-TS  
No. 2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-198395

(43) 公開日 平成10年(1998) 7月31日

(51) Int.Cl.<sup>6</sup>

G 1 0 L 3/00

5/06

識別記号

5 3 5

5 3 1

5 6 1

F I

G 1 0 L 3/00

5/06

5 3 5

5 3 1 F

5 6 1 G

D

審査請求 有 請求項の数 5 O L (全 14 頁)

(21) 出願番号

特願平9-2733

(22) 出願日

平成9年(1997) 1月10日

(71) 出願人 593118597

株式会社エイ・ティ・アール音声翻訳通信  
研究所

・京都府相楽郡精華町大字乾谷小字三平谷5  
番地

(72) 発明者 政瀧 浩和

京都府相楽郡精華町大字乾谷小字三平谷5  
番地 株式会社エイ・ティ・アール音声翻  
訳通信研究所内

(72) 発明者 匂坂 芳典

京都府相楽郡精華町大字乾谷小字三平谷5  
番地 株式会社エイ・ティ・アール音声翻  
訳通信研究所内

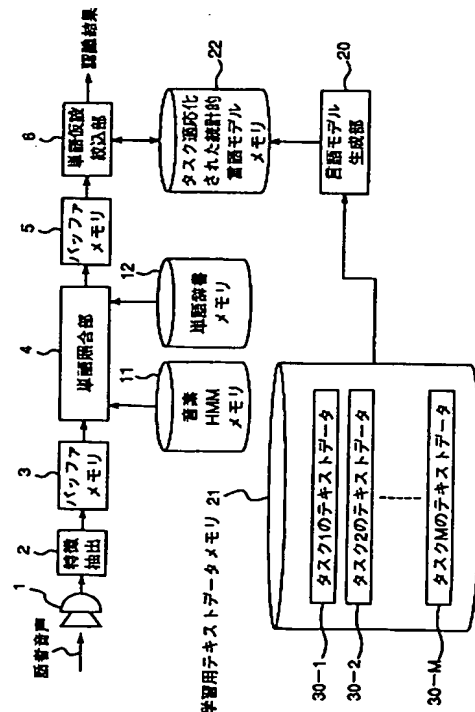
(74) 代理人 弁理士 青山 稔 (外2名)

(54) 【発明の名称】 統計的言語モデル生成装置及び音声認識装置

(57) 【要約】

【課題】 遷移確率の予測精度及び信頼性を改善可能な統計的言語モデルを生成し、統計的言語モデルを用いてより高い音声認識率で音声認識する。

【解決手段】 複数のタスク毎の学習用テキストデータに基づいて最尤推定法を用いて各タスク毎に単語 N-gram の遷移確率を演算し、当該遷移確率の出現分布を事前知識の確率分布と仮定し、各タスク毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する。次いで、演算された事前知識の確率分布のパラメータと、学習用テキストデータ内の特定タスクのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて各タスク毎の単語 N-gram の遷移確率を計算することにより統計的言語モデルを生成する。また、当該統計的言語モデルを用いて音声認識する。



## 【特許請求の範囲】

【請求項1】 所定の複数のタスク毎の学習用テキストデータに基づいて最尤推定法を用いて各タスク毎に単語のN-gram (Nは2以上の自然数である。)の遷移確率を演算する第1の演算手段と、

上記第1の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率の出現分布を事前知識の所定の確率分布と仮定し、各タスク毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第2の演算手段と、

上記第2の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータ内の特定タスクのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて、各タスク毎の単語のN-gramの遷移確率を計算することにより、各タスク毎の単語のN-gramの遷移確率を含む統計的言語モデルを生成する第3の演算手段とを備えたことを特徴とする統計的言語モデル生成装置。

【請求項2】 上記確率分布は、ベータ分布又はディリクレ分布であることを特徴とする請求項1記載の統計的言語モデル生成装置。

【請求項3】 上記統計的言語モデル生成装置はさらに、上記第3の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率に基づいて、所定の平滑化処理を実行し、処理後の各タスク毎の単語のN-gramの遷移確率を含む統計的言語モデルを生成する第4の演算手段を備えたことを特徴とする請求項1又は2記載の統計的言語モデル生成装置。

【請求項4】 入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して音声認識することを特徴とする音声認識装置。

【請求項5】 入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた音声認識装置において、

上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする音声認識装置。

【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、学習用テキストデータに基づいて統計的言語モデルを生成する統計的言語モデル生成装置、及び上記統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識する音声認識装置に関する。

## 【0002】

【従来の技術】 近年、連続音声認識装置において、その性能を高めるために言語モデルを用いる方法が研究されている。これは、言語モデルを用いて、次単語を予測し探索空間を削減することにより、認識率の向上および計算時間の削減の効果を狙ったものである。最近盛んに用いられている言語モデルとしてN-gram (N-グラム; ここで、Nは2以上の自然数である。)がある。これは、大規模なテキストデータを学習し、直前のN-1個の単語から次の単語への遷移確率を統計的に与えるものである。複数L個の単語列 $w_1^L = w_1, w_2, \dots, w_L$ の生成確率 $P(w_1^L)$ は次式で表される。

## 【0003】

【数1】

$$P(w_1^L) = \prod_{t=1}^L P(w_t | w_{1:t-1})$$

【0004】 ここで、 $w_t$ は単語列 $w_1^L$ のうちt番目の1つの単語を表し、 $w_{1:t-1}$ は1番目からj番目の単語列を表す。上記数1において、確率 $P(w_t | w_{1:t-1})$ は、N個の単語からなる単語列 $w_{1:t-1}$ が発声された後に単語 $w_t$ が発声される確率であり、以下同様に、確率 $P(A | B)$ は単語又は単語列Bが発声された後に単語Aが発声される確率を意味する。また、数1における「 $\prod$ 」は $t=1$ からLまでの確率 $P(w_t | w_{1:t-1})$ の積を意味し、以下同様である。

【0005】 ところで、近年、上記統計的言語モデルN-gramを用いて連続音声認識の性能を向上させる手法が盛んに提案されている(例えば、従来技術文献1「L. R. Bahl et al., "A Maximum Likelihood Approach to Continuous Speech Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 179-190, 1983年」及び従来技術文献2「清水ほか, "単語グラフを用いた自由発話音声認識", 電子情報通信学会技術報告, SP95-88, pp. 49-54, 平成7年」参照。)

【0006】 しかしながら、N-gramはパラメータ数が多く、それぞれの値を正確に求めるためには、莫大な量のテキストデータが必要とされる。この問題を解決する方法として、学習用テキストデータに出現しない単語遷移に対しても遷移確率を与える平滑化の手法(例え

ば、従来技術文献3「F. Jelinek et al., "Interpolated estimation of Markov Source Parameters from Sparse Data", Proceedings of Workshop Pattern Recognition in Practice, pp. 381-387, 1980年」、従来技術文献4「S. M. Katz et al., "Estimation of Probabilities from Sparse Data for the Language model Component of a Speech Recognizer", IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 400-401, 1987年」及び従来技術文献5「川端ほか, "二項事後分布に基づくN-gram言語モデルのBack-off平滑化", 電子情報通信学会技術報告, SP95-93, pp. 1-6, 平成7年」参照。)や、クラス分類、可変長N-gram等パラメータの数を減少させる手法(例えば、従来技術文献6「P. F. Brown et al., "Class-Based n-gram models of natural language", Computational Linguistics, Vol. 18, No. 4, pp. 467-479, 1992年」、従来技術文献7「T. R. Niesler et al., "A Variable-Length Category-Based N-gram Language Model", Proceedings of ICASSP '96, Vol. 1, pp. 164-167, 1996年」及び従来技術文献8「政瀧ほか, "連続音声認識のための可変長連鎖統計言語モデル", 電子情報通信学会技術報告, SP95-73, pp. 1-6, 平成7年」参照。)等が数多く提案されている。しかしながら、これらの手法を用いても、精度の良い言語モデルを構築するためには、相当量のデータを用いる必要があると考えられる。

【0007】現在、実用化に向けて研究が行われている連続音声認識システムは、ホテル予約・スケジューリング等、タスクを限定し、システムのパフォーマンスの向上させている物が多い。しかしながら、タスク毎に大量の言語データを集めるのは困難である。特に、日本語の場合は、英語等のように単語の区切りが明確ではなく、通常人間が手作業で単語の切り出し・形態素解析の作業を行うため、大量のデータを集めるのはさらに困難である。しかしながら、データ量を増やすために、他のタスクのデータを用いる場合、言語的特徴はタスク毎に異なるため、単純にデータを混合しても目的のタスク特有の言語特徴を効果的に表現することはできないと考えられる。

【0008】これらの問題を解決する手段として、言語モデルのタスク適応を考えられている。すなわち、目的のタスク以外のデータも含めた大量のデータを学習することによりデータ量の問題を解決し、得られたモデルの言語特徴を目的のタスクに適応させる方法である。タスク適応の手法として、従来提案されているものには大量のデータで作成したN-gramと、目的タスクのデータで作成したN-gramとを重みづけにより混合する方法(例えば、従来技術文献9「伊藤ほか, "小量テキストによる言語モデルのタスク適応の検討", 日本音響学会講演論文集, 1-3-19, pp. 37-38, 平成8年9月」及び従来技術文献10, 「坂本ほか, "音声認識における統計的言語モデルの選択の効果", 日本音響学会講演論文集, 1-Q-24, pp. 157-158, 平成6年3月」参照。)がある。

【0009】例えば、従来技術文献9の手法を用いた従来例のタスク適応化処理は、小量テキストに重みをかけて大量テキストと混合するものであり、次の手順によって言語モデルを作成する。

(a) 小量テキストを、重み付きで大量テキストに加える。重み係数を $w$ とすると、小量テキスト中で $m$ 回出現した単語は、大量テキスト中で $(w \cdot m)$ 回出現した単語と同等に扱われることになる。

(b) このようにしてできたテキストの中で、出現頻度が一定回数未満の単語を、未知語を表す記号に置き換える。すなわち、語彙の制限を行う。

(c) テキストから統計を取り、N-gramモデルを構築する。

【0010】

【発明が解決しようとする課題】しかしながら、従来例のタスク適応化処理においては、重み係数 $w$ は1つのみしか使用していないので、言語モデルにおける遷移確率の予測精度はいまだ比較的低く、当該言語モデルを用いて音声認識をしたときの音声認識率は比較的低いという問題点があった。

【0011】本発明の目的は以上の問題点を解決し、従来例に比較して遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを生成することができる統計的言語モデル生成装置、及び、当該統計的言語モデルを用いて従来例に比較して高い音声認識率で音声認識することができる音声認識装置を提供することにある。

【0012】

【課題を解決するための手段】本発明に係る請求項1記載の統計的言語モデル生成装置は、所定の複数のタスク毎の学習用テキストデータに基づいて最尤推定法を用いて各タスク毎に単語のN-gram( $N$ は2以上の自然数である。)の遷移確率を演算する第1の演算手段と、上記第1の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率の出現分布を事前知識の所

定の確率分布と仮定し、各タスク毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第2の演算手段と、上記第2の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータ内の特定タスクのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて、各タスク毎の単語のN-gramの遷移確率を計算することにより、各タスク毎の単語のN-gramの遷移確率を含む統計的言語モデルを生成する第3の演算手段とを備えたことを特徴とする。

【0013】また、請求項2記載の統計的言語モデル生成装置は、請求項1記載の統計的言語モデル生成装置において、上記確率分布は、ベータ分布又はディリクレ分布であることを特徴とする。

【0014】さらに、請求項3記載の統計的言語モデル生成装置は、請求項1又は2記載の統計的言語モデル生成装置において、さらに、上記第3の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率に基づいて、所定の平滑化処理を実行し、処理後の各タスク毎の単語のN-gramの遷移確率を含む統計的言語モデルを生成する第4の演算手段を備えたことを特徴とする。

【0015】本発明に係る請求項4記載の音声認識装置は、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して音声認識することを特徴とする。

【0016】本発明に係る請求項5記載の音声認識装置は、入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする。

【0017】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。図1に本発明に係る一実施形態の連続音声認識装置のブロック図を示す。本実施形態の連続音声認識装置は、公知のワンパス・ビタビ復号化法を用いて、入力される発声音声文の音声信号の特徴パラメータに基づいて上記発声音声文の単語仮説を

検出し尤度を計算して出力する単語照合部4を備えた連続音声認識装置において、単語照合部4からバッファメモリ5を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、タスク適応化された統計的言語モデルメモリ（以下、統計的言語モデルメモリという。）22内のタスク適応化された統計的言語モデルを参照して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行う単語仮説絞り部6を備えたことを特徴とする。

【0018】ここで用いる統計的言語モデルは、それぞれ単語列からなる各タスク $m$ のテキストデータ30-m ( $m=1, 2, \dots, M$ )を区分されて備えた学習用テキストデータに基づいて言語モデル生成部20により生成されたものであって、統計的言語モデル22は、各タスク毎にタスク適応化された各タスク毎の単語間のN-gram ( $N=2, 3$ )を含む。本発明に係る実施形態においては、当該統計的言語モデルを生成するときに、MAP (Maximum A-posteriori Probability: 最大事後確率) 推定法を用いた処理を用いる。この処理においては、図3に示すように、不特定タスクのテキストデータ30-1乃至30-Mに基づいて生成されるN-gramを含む統計的言語モデル31を事前知識とする一方、目的の特定タスク $i$ のテキストデータ30- $i$ を事後知識としてMAP推定法を用いて、データ量に応じて安定性の高いパラメータ推定を行った後、Back-off平滑化法による平滑化処理を実行することによりタスク適応化されたN-gramを含む統計的言語モデルを得ることを特徴とする。ここで、目的の特定タスク $i$ のテキストデータ30- $i$ は、不特定タスクのテキストデータ30-m ( $m=1, 2, \dots, M$ )のうちの1つの特定タスクのテキストデータである。

【0019】まず、MAP推定法による遷移確率の算出について述べる。通常、N-gramの遷移確率は、ML (Maximum Likelihood; 最尤) 推定法により算出される。最尤推定法は、観測したサンプル値（すなわち、テキストデータの単語） $x$ に対して、遷移確率 $p$ が決まったときに単語 $x$ となる尤度関数 $f(x|p)$ を最大にさせる値として、N-gram遷移確率 $p_{ML}$ が次式で定められる。

【0020】

【数2】

$$p_{ML} = \underset{p}{\operatorname{argmax}} f(x|p)$$

【0021】ここで、関数 $\operatorname{argmax}$ は、 $p$ を変化したときに関数の引数が最大となるときの $p$ の値を示す関数である。議論を簡単にするため、単語 $w_k$ から $w_l$ へのbigramの遷移確率 $p(w_l|w_k)$ について考え、

また、以下のような変数の定義を行う。

(a)  $N$ : 学習用テキストデータ中の単語  $w_k$  の出現回数。

(b)  $n$ : 学習用テキストデータ中の単語列  $w_k w_l$  の出現回数。

(c)  $p$ : 単語  $w_k$  から  $w_l$  への遷移確率。

【0022】このとき尤度関数  $f(x|p)$  は、単語  $w_k$  が  $N$  回観測され、次に単語  $w_l$  に続く回数が  $n$  回で、それ以外の単語に続く回数が  $(N-n)$  回であるから、次式を得ることができる。

【0023】

$$【数3】 f(p) = p^n (1-p)^{N-n}$$

【0024】 $f(p)$  の最大化条件  $d \log f(p) / dp = 0$  を解くことにより、 $N$ -gram の遷移確率は次式のように計算される。

【0025】

$$【数4】 p_{ML} = n/N$$

【0026】従って、もし単語列  $w_k w_l$  が観測データ上に出現しない場合、 $n=0$  であるから、遷移確率は0と推定されてしまう。これに対して、MAP (Maximum A-posteriori Probability; 最大事後確率) 推定においては、最尤推定法を用いて、観測したサンプル値  $x$  に対して、遷移確率  $p$  が決定される事後確率関数  $h(p|x)$  を最大化する値として、 $N$ -gram の遷移確率が次式で求められる。

【0027】

【数5】

$$p_{MAP} = \underset{p}{\operatorname{argmax}} h(p|x)$$

【0028】ここで、Bayes 則を用いると、上記数5は次式のように変形される。

【0029】

【数6】

$$p_{MAP} = \underset{p}{\operatorname{argmax}} f(x|p) g(p)$$

【0030】ここで、 $g(p)$  は、不特定タスクのテキストデータ  $30-1$  乃至  $30-M$  に基づいて予め決定さ

$$\begin{aligned} p_{MAP} &= \underset{p}{\operatorname{argmax}} \{ p^n (1-p)^{N-n} a p^{\alpha-1} (1-p)^{\beta-1} \} \\ &\equiv \underset{p}{\operatorname{argmax}} L(p) \end{aligned}$$

【0034】ここで、関数  $L(p)$  が最大となるための条件  $d \log L(p) / dp = 0$  を  $p$  について解くと、単語の bigram の遷移確率  $p_{MAP}$  は次式のように求まる。

【0035】

$$【数8】 p_{MAP} = (n + \alpha - 1) / (N + \alpha + \beta - 2)$$

【0036】ここで、パラメータ  $\alpha$  及び  $\beta$  は、事前分布であるベータ分布のパラメータであるが、これらは、次

れる、 $N$ -gram の遷移確率  $p$  の事前分布である。すなわち、MAP 推定法を用いると、 $N$ -gram の遷移確率はある事前知識より得られる分布  $g(p)$  に従う変数とし、この事前分布と実際に観測されたサンプル値とを用いて、実際の遷移確率が推定される。このため、観測データで出現しない単語遷移に対しても、事前知識により0でない遷移確率を与えることができる。

【0031】次いで、bigram を例にとり、MAP 推定法により  $N$ -gram の遷移確率を求める方法を示す。ただし、変数の定義は上述と同じものを用いる。まず、遷移確率  $p$  の事前分布としてベータ分布 ( $a p^{\alpha-1} (1-p)^{\beta-1}$ 、ここで、 $\alpha$  及び  $\beta$  はベータ分布の正のパラメータであり、 $a$  は正規化のための正の定数である。) を用いる。なお、 $0 \leq p \leq 1$  である。ベータ分布を用いる理由は次の2点である。

(a) ベータ分布は2項分布の自然共役事前分布で、MAP 推定によるパラメータの解が求まりやすい。

(b) ベータ分布のパラメータ  $\alpha$ 、 $\beta$  を変化させることにより、様々な形状の分布を表すことができる。

ここで、ベータ分布は、連続変数の確率分布の一種であり、ガンマ関数をもとにして構成されるベータ関数が表式に含まれる。なお、本実施形態においては、ベータ分布を用いるが、本発明はこれに限らず、ベータ分布に代えて、ディリクレ分布を用いてもよい。ディリクレ分布は、ベータ分布を多変量分布に拡張したものであり、多項分布の自然共役事前分布である。

【0032】上記数6のMAP 推定法の定義に従うと、遷移確率  $p_{MAP}$  は、尤度関数  $f(p)$  と事前分布  $g(p)$  とを用いて次式のように求められる。

【0033】

【数7】

式のように求めることができる。なお、ベータ分布の平均  $\mu$  及び分散  $\sigma^2$  は以下の式となることが知られている (例えば、従来技術文献5参照。)

【0037】

$$【数9】 \mu = \alpha / (\alpha + \beta)$$

$$\sigma^2 = (\alpha \beta) / \{ (\alpha + \beta)^2 (\alpha + \beta + 1) \}$$

【0038】これらの式を  $\alpha$ 、 $\alpha + \beta$  について解くと、次式が得られる。

【0039】

【数10】  $\alpha = \{\mu^2(1-\mu)\} / \sigma^2 - \mu$  $\alpha + \beta = \{\mu(1-\mu)\} / \sigma^2 - 1$ 

【0040】以上より、観測テキストデータから頻度を計算することにより得られるパラメータ $N$ 、 $n$ 、及び事前分布の平均 $\mu$ 及び分散 $\sigma^2$ により、上記数8及び数10を用いて、単語のbigramの遷移確率を求めることができる。

【0041】これまでの議論は、単語のbigramについてのみの議論であったが、一般に、MAP推定法によるN-gramの遷移確率 $p(w_n | w_1^{n-1})$ は、直前の単語 $w_k$ を直前の単語列 $w_1^{n-1}$ と置き換え、パラメータ $N$ 及び $n$ を次のように定義すれば、同じ議論が通用することは明らかである。

(a)  $N$ : 学習用テキストデータ中の単語列 $w_1^{n-1}$ の出現回数 ( $c(w_1^{n-1})$ )、すなわち、処理対象の単語列の直前の単語列の出現回数である。

(b)  $n$ : 学習用テキストデータ中の単語列 $w_1^n$ の出現回数 ( $c(w_1^n)$ )、すなわち、処理対象の単語列の出

$$\mu = \sum_i c_i(w_n) p_i(w_n | w_1^{n-1}) / \sum_i c_i(w_1^{n-1})$$

【数12】

$$\sigma^2 = \sum_i c_i(w_1^{n-1}) p_i(w_n | w_1^{n-1})^2 / \sum_i c_i(w_1^{n-1}) - \mu^2$$

【0044】ここで、 $c_i(w_1^{n-1})$ はタスク $i$ において単語列 $w_1^{n-1}$ の出現頻度であり、 $p_i(w_n | w_1^{n-1})$ はタスク $i$ における単語列 $w_1^{n-1}$ から $w_n$ への遷移確率である。また、事後知識を目的のタスクのテキストデータとすると、前述のパラメータ $N$ 及び $n$ は次のように表される。

(a)  $N$ : 目的の特定タスク $i$ のテキストデータ30-i中の単語列 $w_1^{n-1}$ の出現頻度、すなわち、処理対象の単語列の直前の単語列の出現回数である。

(b)  $n$ : 目的の特定タスク $i$ のテキストデータ30-i中の単語列 $w_1^n$ の出現頻度、すなわち、処理対象の単語列の出現回数である。

以上の加重平均 $\mu$ 、加重分散 $\sigma^2$ 、パラメータ $n$ 及び $N$ を上記の数7及び数9に代入することにより、MAP推定法によるタスク適応後のN-gram遷移確率が得られる。

【0045】さらに、Back-off平滑化法による遷移確率の平滑化について述べる。上記でMAP推定法によるタスク適応の基本原則を述べたが、実際に言語モデルとして使用するには、2つの問題がある。1つは、平滑化の問題である。不特定タスクの大量のテキストデータを用いても、出現しない単語列が存在し、MAP推定法を用いても、N-gramの遷移確率が0となってしまう。従って、平滑化処理によりテキストに出現しな

$$P_s(w_n | w_1^{n-1}) = P_h(w_n | w_1^{n-1}), \quad c_i(w_1^{n-1}) > 0 \text{ のとき}$$

現回数である。

【0042】次いで、MAP推定法を用いたタスク適応化処理について述べる。上述のMAP推定法によるN-gramをタスク適応化に応用するために、図3に示すように、複数のタスクより構成される大量のテキストデータから作成される不特定タスクのテキストデータ30-mに基づくN-gramを事前知識とし、目的の特定タスク $i$ のテキストデータ30-iを事後知識とみなす。不特定タスクのN-gramを事前知識とみなしたとき、その事前分布は、各タスクにおけるN-gram遷移確率の分布と考えることができる。ただし、各タスクにおけるN-gram遷移確率は最尤推定法により求められる。この事前分布をベータ分布と仮定し、図4に示すように、MAP推定法の前分布として用いる。このとき、事前分布の加重平均 $\mu$ 、及び加重分散 $\sigma^2$ は次式で求められる。

【0043】

【数11】

い単語組に対しても、0でない遷移確率を与える必要がある。もう1つの問題は、本発明に係るタスク適応化処理は、全ての遷移確率を独立に求める手法であるため、遷移確率の和が1になるとは限らない。連続音声認識等に適用する際は、問題とはならないが、パープレキシティで評価する際は、1に正規化されていないと、正しい評価ができない。従って、近年盛んに用いられているBack-Off平滑化法（例えば、従来技術文献4参照。）を拡張して、これらの問題を解決する方法を述べる。

【0046】単語列 $w_1^n$ が不特定タスクのテキストデータ30-mに含まれる場合は、上記のタスク適応化処理により、遷移確率 $p_{MAP}(w_n | w_1^{n-1})$ を求め、チューリング(Turing)推定法により、確率 $p_{MAP}(w_n | w_1^{n-1})$ を軽減する。ただし、軽減係数は不特定タスクのテキストデータ30-mの頻度 ( $c_i(w_1^n)$ ) を用いて計算する。当該軽減により生じた確率の余剰分を $w_1^n$ が不特定タスクのテキストデータ30-mに含まれない単語連鎖に対して、 $(n-1)$ -gramの遷移確率に比例して配分する。以上をまとめると、タスク適応化されたN-gramの平滑化後の遷移確率 $P_s(w_n | w_1^{n-1})$ は次式で表される。

【0047】

【数13】

$$= \alpha (w_1^{n-1}) P s (w_0 | w_2^{n-1}), c_1 (w_1^{n-1}) = 0, c_1 (w_2^{n-1}) > 0 \text{ のとき}$$

$$= P s (w_0 | w_2^{n-1}), c_1 (w_1^{n-1}) = 0, c_1 (w_2^{n-1}) = 0 \text{ のとき}$$

【0048】上記の数12において、Phはタスク適応により得られる確率に軽減係数をかけたものであり、次式で与えられる。

$$\begin{aligned} Ph (w_0 | w_1^{n-1}) \\ = \{ c_1 (w_1^n) + 1 \} / \{ c_1 (w_1^n) \} \\ \times \{ n c_1 (w_1^n) + 1 \} / \{ n c_1 (w_1^n) \} \cdot p_{MAP} (w_1^n) \end{aligned}$$

【0050】ここで、 $n_c$ は、不特定タスクのテキストデータ30-m中にc回出現する単語列の種類数（異なり）であり、また、数13で、 $\alpha (w_1^{n-1})$ は正規化のための係数であり、次のように求められる。

【0051】

$$[\text{数15}] \alpha (w_1^{n-1}) = A a / A b$$

ここで、

$$A a \equiv 1 - \sum_{w_1: c_1 (w_1^n) > 0} Ph (w_0 | w_1^{n-1})$$

$$A b \equiv 1 - \sum_{w_1: c_1 (w_2^n) > 0} Ph (w_0 | w_1^{n-1})$$

【0052】以上のBack-off平滑化法を応用した手法を用いることにより、学習データ上に出現しない単語連鎖に対しても確率値を与えることができる。また、遷移確率 $p_{MAP}$ が正規化されていなくても、上記数15において $\alpha$ を求める際に正規化の計算を行うため、遷移確率の和は自動的に1になる。

【0053】図5は、言語モデル生成部20によって実行される言語モデル生成処理の詳細を示すフローチャートであり、以下、図5を参照して当該処理について説明する。なお、以下の各処理は各単語列毎に実行される。まず、ステップS1では、各タスクmのテキストデータ30-mを記憶する学習用テキストデータメモリ21から各タスクmのテキストデータ30-mを読み出す。次いで、ステップS2において、ステップS2において、読み出した各タスクのテキストデータ30-mに基づいて最尤推定法を用いて各タスク毎に単語のbigramの遷移確率を数3を用いて計算する。そして、ステップS3において、各タスク毎の単語bigramの遷移確率の出現分布をベータ分布と仮定し、ベータ分布の加重平均 $\mu$ 及び加重分散 $\sigma^2$ を数10及び数11を用いて計算した後、これらに基づいて数9を用いてベータ分布のパラメータ $\alpha$ 及び $\beta$ を計算する。さらに、ステップS4において、事前知識のパラメータ $\alpha$ 及び $\beta$ と、特定タスクiのテキストデータ30-iの事後知識のパラメータN及びnとに基づいて、数7を用いて各タスク毎の単語bigramの遷移確率pを計算する。そして、ステップS5において、各タスク毎の単語bigramの遷移確率pに基づいて、Back-off平滑化処理を実行し、処理後の各タスク毎の単語bigramの遷移確率を含む統計的言語モデルを生成して、統計的言語モデルメモリ22に記憶する。以上で、言語モデル生成処理を

【0049】

【数14】

終了する。

【0054】次いで、図1に示す連続音声認識装置の構成及び動作について説明する。図1において、単語照合部4に接続された音素隠れマルコフモデル（以下、隠れマルコフモデルをHMMという。）メモリ11内の音素HMMは、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

(a) 状態番号、(b) 受理可能なコンテキストクラス、(c) 先行状態、及び後続状態のリスト、(d) 出力確率密度分布のパラメータ、及び(e) 自己遷移確率及び後続状態への遷移確率。なお、本実施形態において用いる音素HMMは、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して生成する。ここで、出力確率密度関数は34次元の対角共分散行列をもつ混合ガウス分布である。また、単語照合部4に接続された単語辞書メモリ12内の単語辞書は、音素HMMメモリ11内の音素HMMの各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【0055】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 $\Delta$ 対数パワー及び16次 $\Delta$ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して単語照合部4に入力される。

【0056】単語照合部4は、ワンパス・ビタビ復号化法を用いて、バッファメモリ3を介して入力される特徴パラメータのデータに基づいて、音素HMM11と単語辞書12とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部4は、各時刻の各HMMの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素HMM11及び単語辞書12とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部4は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ5を介して単語仮説校正部6に出力する。

【0057】単語仮説校正部6は、単語照合部4からバッファメモリ5を介して出力される単語仮説に基づいて

て、統計的言語モデルメモリ22内のタスク適応化された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として出力する。なお、タスク適応化された統計的言語モデルは、各タスク毎に1つの統計的言語モデルを備え、単語仮説絞り込みは、音声認識しようとするタスクに対応する統計的言語モデルを選択的に参照する。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びをいう。

【0058】例えば、図2に示すように、 $(i-1)$ 番目の単語 $W_{i-1}$ の次に、音素列 $a_1, a_2, \dots, a_n$ からなる $i$ 番目の単語 $W_i$ がくるときに、単語 $W_{i-1}$ の単語仮説として6つの仮説 $W_a, W_b, W_c, W_d, W_e, W_f$ が存在している。ここで、前者3つの単語仮説 $W_a, W_b, W_c$ の最終音素は $/x/$ であるとし、後者3つの単語仮説 $W_d, W_e, W_f$ の最終音素は $/y/$ であるとする。終了時刻 $t_e$ と先頭音素環境が等しい仮説（図2では先頭音素環境が $"x/a_1/a_2"$ である上から3つの単語仮説）のうち総尤度が最も高い仮説（例えば、図2において1番上の仮説）以外を削除する。なお、上から4番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素が $x$ ではなく $y$ であるので、上から4番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に1つのみ仮説を残す。図2の例では、最終音素 $/x/$ に対して1つの仮説を残し、最終音素 $/y/$ に対して1つの仮説を残す。

【0059】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素

と連続する先行する単語仮説の少なくとも1つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【0060】以上の実施形態において、特徴抽出部2と、単語照合部4と、単語仮説絞り込み部6と、言語モデル生成部20とは、例えば、デジタル電子計算機などのコンピュータで構成され、バッファメモリ3、5と、音素HMMメモリ11と、単語辞書メモリ12と、学習用テキストデータメモリ21と、統計的言語モデルメモリ22とは、例えばハードディスクメモリなどの記憶装置で構成される。

【0061】以上実施形態においては、単語照合部4と単語仮説絞り込み部6とを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素HMM11を参照する音素照合部と、例えばOne Pass DPアルゴリズムを用いて統計的言語モデルを参照して単語の音声認識を行う音声認識部とで構成してもよい。

【0062】

【実施例】本発明者は、本実施形態で用いるタスク適応化された統計的言語モデルの性能を確認するため、評価実験を行った。実験用いたデータは、本特許出願人が所有する自然発話データベース（例えば、従来技術文献11「T. Morimoto et al., "A Speech and Language Database for Speech Translation Research", ICSLP, pp. 1791-1794, 1994年」参照。）であり、現在1,098会話、449,070単語（のべ）、6,797（異なり）単語で構成されている。また、このデータベースは表1に示すように、15タスクから構成されており、各タスクの単語数を表2に示す。これらのデータのうち、約1/4の会話をランダムに選んでテストセットとして、残りの会話を学習セットとして使用した。ただし、各タスクから最低でも1会話はテストセットとして選択している。

【0063】

【表1】

タスク番号	会話数	内容
1	491	ホテルのサービス
2	351	ホテルの部屋の予約
3	50	観光バス・旅行パックの問い合わせ
4	36	ホテルの会議室の相談・予約
5	28	交通手段の問い合わせ
6	24	ホテルの部屋の相談
7	22	飛行機のフライトの予約
8	22	定期バス・列車の切符の問い合わせ



9	20	レンタカーの問い合わせ
10	14	演劇・コンサートのチケットの予約
11	12	レストランの予約
12	8	トラブル・忘れ物
13	8	料理の注文
14	8	道案内
15	4	ショッピング

【0064】

【表2】

タスクの単語数

タスク番号	学習セット	テストセット
1	136, 175	42, 698
2	118, 124	38, 697
3	19, 471	6, 610
4	15, 302	5, 075
5	10, 791	2, 983
6	8, 802	2, 999
7	8, 617	2, 722
8	8, 537	2, 193
9	8, 567	2, 528
10	5, 036	1, 608
11	5, 326	1, 439
12	3, 578	1, 165
13	2, 378	1, 075
14	2, 572	908
15	1, 750	509

【0065】ここで、本発明に係る統計的言語モデルを評価するために、パープレキシティを用いる。例えば、複数  $n$  個の単語からなる長い単語列  $w_1^n = w_1 w_2 \dots w_n$  があるときのエントロピー  $H(n)$  は次式で表される。

実験条件

分析条件 サンプル周波数：12 KHz,  
ハミング窓：20 ms,  
フレーム周期：10 ms

使用パラメータ 16次LPCケプストラム+16次 $\Delta$ ケプストラム  
+logパワー+ $\Delta$ logパワー

音響モデル HM網の男女別不特定話者モデル  
400状態、5混合

【0071】表3において、HM網の男女別不特定話者モデルについては、従来技術文献12「小坂ほか，“話者混合SSSによる不特定話者音声認識”，日本音響学会講演論文集，2-5-9，pp. 135-136，平

【0066】

【数16】

$$H(n) = -(1/n) \cdot \log_2 P(w_1^n)$$

【0067】ここで、 $P(w_1^n)$  は単語列  $w_1^n$  の生成確率であり、パープレキシティ  $PP(n)$  は次式で表される。

【0068】

$$PP(n) = 2^{H(n)}$$

【0069】当該実験における統計的言語モデルとしては、次の3種類のモデルを考える。

(a) 不特定タスクモデル：全タスクのテキストで作成した  $N\text{-gram}$ 。

(b) 特定タスクモデル：各タスクのテキストのみで作成した  $N\text{-gram}$ 。

(c) タスク適応モデル：不特定タスクモデルを各タスクに適応した  $N\text{-gram}$ 。

これらの統計的言語モデルをタスク毎に、単語の  $bigram$  及び単語の  $trigram$  で作成した。そして、本発明者は、本実施形態のタスク適応化された統計的言語モデルを図1の連続音声認識装置に適用し、当該統計的言語モデルの効果を確認した。音素認識の実験条件を表3に示す。また、音響をパラメータもあわせて表3に示す。

【0070】

【表3】

成4年」に開示されている。この実験では、単語グラフを用いた連続音声認識法を用いて音響モデルおよび言語モデルを連続音声認識装置に適用した。さらに、各モデル及び各タスク毎のテストセットパープレキシティ値を

表4に示す。表4において、不特定タスクモデルはすべてのタスクの学習セットを用いて学習される一方、特定タスクモデルとタスク適応化モデルは各対応するタスクの学習セットを用いて学習される。そして、パーブレキ各モデルのタスク別パーブレキシティ

シティ値は各タスクのテストセットを入力データとして用いたときに得られた値である。

【0072】

【表4】

タスク番号	不特定タスクモデル		特定タスクモデル		タスク適応化モデル	
	Bigram	Trigram	Bigram	Trigram	Bigram	Trigram
1	23.168	17.948	22.923	18.260	22.085	17.515
2	14.837	10.071	13.842	9.941	13.402	9.612
3	26.523	17.383	23.910	17.196	20.684	14.705
4	31.270	24.693	38.164	32.811	29.280	24.470
5	24.164	16.544	21.774	16.574	18.328	13.656
6	17.122	11.192	14.661	11.350	12.540	9.127
7	21.106	14.181	18.358	14.656	15.274	11.383
8	21.134	14.288	14.077	11.177	13.351	10.523
9	25.149	18.154	25.897	20.743	20.443	16.097
10	16.582	10.820	14.060	10.931	11.368	8.148
11	12.970	8.867	12.261	9.611	9.564	6.935
12	32.921	19.402	25.232	18.385	19.921	13.399
13	30.294	22.416	32.757	31.567	21.541	19.338
14	35.490	27.108	45.853	41.285	28.155	23.707
15	44.088	34.214	47.324	44.573	31.854	27.896
平均	25.121	17.819	24.740	20.604	19.186	15.101

【0073】表4から明らかなように、タスク適応化モデルのパーブレキシティは、不特定タスクモデルと比較して、平均で24% (bigram), 15% (trigram) 低くなっている。特定タスクモデルと比較しても、平均で22% (bigram), 27% (trigram) 低い。さらに、タスク適応化モデルのパーブレキシティは、全てのタスクでbigram, trigramの両方の場合において、不特定タスクモデル及び特定モデルのいずれよりも低い。従って、本実施形態の方法により、N-gramのタスク適応化が効果的に行われていることが示せた。

【0074】不特定タスクモデルと特定タスクモデルのパーブレキシティを比較すると、bigramでは、特定モデルのパーブレキシティの方が不特定モデルよりも低い値を示す場合が多いが、trigramでは、不特定タスクモデルの方が特定タスクモデルよりも低い場合が多い。これは、単語bigramでは、学習のスパース性(密度がうすい、つまり疎であること。)が低いいため、特徴を表すことのできる特定タスクモデルの方が有利であるが、trigramでは、学習がよりスパースであるため、特定タスクの少ない量のテキストデータでは、信頼できるパラメータ推定が行われていないことが

原因と考えられる。従って、タスク適応を行うと、大量のテキストデータを用いたことにより、学習のスパース性が解決でき、さらに、タスク適応化を行うことにより、そのタスクの言語特徴を表現できたものと考えられる。

【0075】テキストデータ量が少ないタスク12及び15等では、タスク適応化によるパーブレキシティの減少が非常に大きい。特にタスク12では、不特定タスクモデルと比較して39% (bigram) 及び30% (trigram)、特定タスクモデルと比較して21% (bigram) 及び27% (trigram) パーブレキシティが減少している。すなわち、目的のタスクのテキストデータが少量しか集まらない場合に、タスク適応化を使用する効果が大きいと言える。ただし、学習量の多いタスク1及び2のタスクにおいては、タスク適応化によるパーブレキシティの減少の割合が小さい。これは、特定タスクモデルにおいては、そのタスクのテキストデータ量が多いため、学習のスパース性がある程度解決できているためと考えられる。

【0076】以上説明したように、本実施形態によれば、MAP推定法を用いることにより、大量のデータから作成されるN-gramをデータ量に応じて目的のタ

スクに効果的に適応化を行う手法を提案した。実験の結果、タスク適応化によるパープレキシティの減少効果が確認され、数千語程度の少量のテキストデータを用いるだけで、適応化前のモデルよりも大幅に精度の良いN-gramが構築できることがわかった。すなわち、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができるとともに、当該統計的言語モデルを用いて音声認識することにより、従来例に比較してより高い音声認識率で連続的に音声認識することができる。

【0077】以上の実施形態において、統計的言語モデルは、N-gramの言語モデルを含むが、ここで、Nは2及び3に限らず、4以上の自然数であってもよい。

【0078】

【発明の効果】以上詳述したように本発明に係る請求項1記載の統計的言語モデル生成装置によれば、所定の複数のタスク毎の学習用テキストデータに基づいて最尤推定法を用いて各タスク毎に単語のN-gram(Nは2以上の自然数である。)の遷移確率を演算する第1の演算手段と、上記第1の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率の出現分布を事前知識の所定の確率分布と仮定し、各タスク毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第2の演算手段と、上記第2の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータ内の特定タスクのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて、各タスク毎の単語のN-gramの遷移確率を計算することにより、各タスク毎の単語のN-gramの遷移確率を含む統計的言語モデルを生成する第3の演算手段とを備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【0079】また、請求項2記載の統計的言語モデル生成装置においては、請求項1記載の統計的言語モデル生成装置において、上記確率分布は、ベータ分布又はディリクレ分布である。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【0080】さらに、請求項3記載の統計的言語モデル生成装置においては、請求項1又は2記載の統計的言語モデル生成装置において、さらに、上記第3の演算手段によって演算された各タスク毎の単語のN-gramの遷移確率に基づいて、所定の平滑化処理を実行し、処理後の各タスク毎の単語のN-gramの遷移確率を含む

統計的言語モデルを生成する第4の演算手段を備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【0081】本発明に係る請求項4記載の音声認識装置においては、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して音声認識する。従って、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを用いて音声認識することにより、従来例に比較してより高い音声認識率で連続的に音声認識することができる。

【0082】本発明に係る請求項5記載の音声認識装置は、入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行う。従って、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを用いて音声認識することにより、従来例に比較してより高い音声認識率で連続的に音声認識することができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施形態である連続音声認識装置のブロック図である。

【図2】 図1の連続音声認識装置における単語仮説絞込部6の処理を示すタイミングチャートである。

【図3】 図1の言語モデル生成部20の処理を示すブロック図である。

【図4】 図1の言語モデル生成部20によって実行されるN-gramのタスク適応化処理における事前分布を示すグラフである。

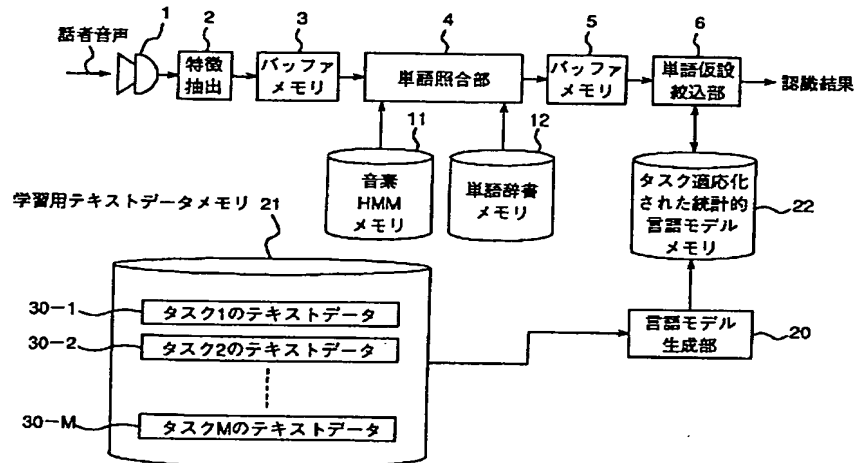
【図5】 図1の言語モデル生成部20によって実行される言語モデル生成処理を示すフローチャートである。

【符号の説明】

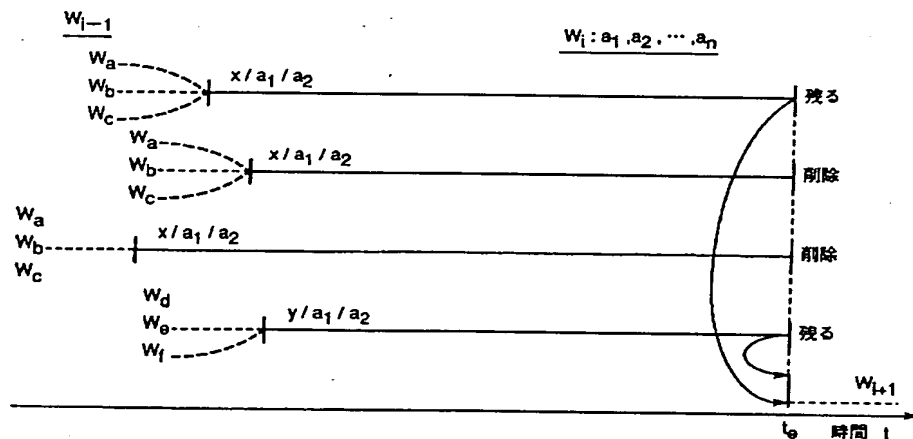
- 1…マイクロホン、
- 2…特徴抽出部、
- 3、5…バッファメモリ、
- 4…単語照合部、
- 6…単語仮説絞込部、
- 11…音素HMMメモリ、
- 12…単語辞書メモリ、

- 20…言語モデル生成部、  
 21…学習用テキストデータメモリ、  
 22…タスク適応化された統計的言語モデルメモリ、  
 30-1乃至30-M…タスクのテキストデータ、  
 31…不特定タスクのN-gramを含む統計的言語モデル、  
 32…MAP推定法を用いたパラメータ推定処理、  
 33…平滑化処理。

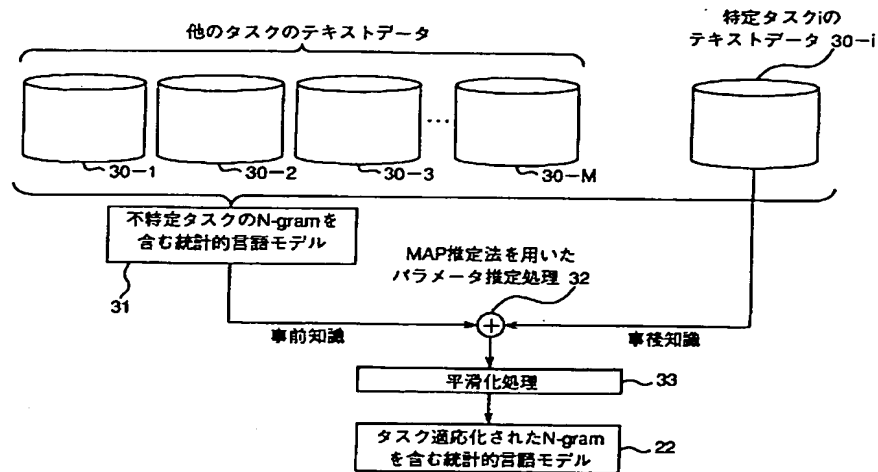
【図1】



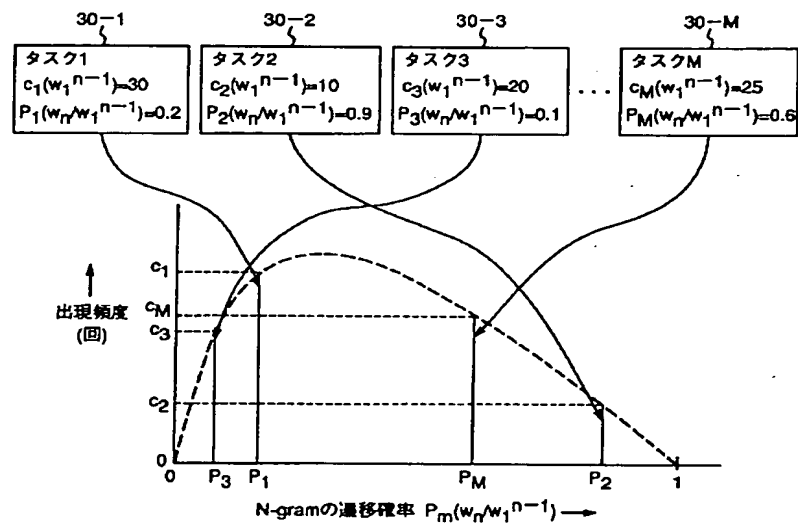
【図2】



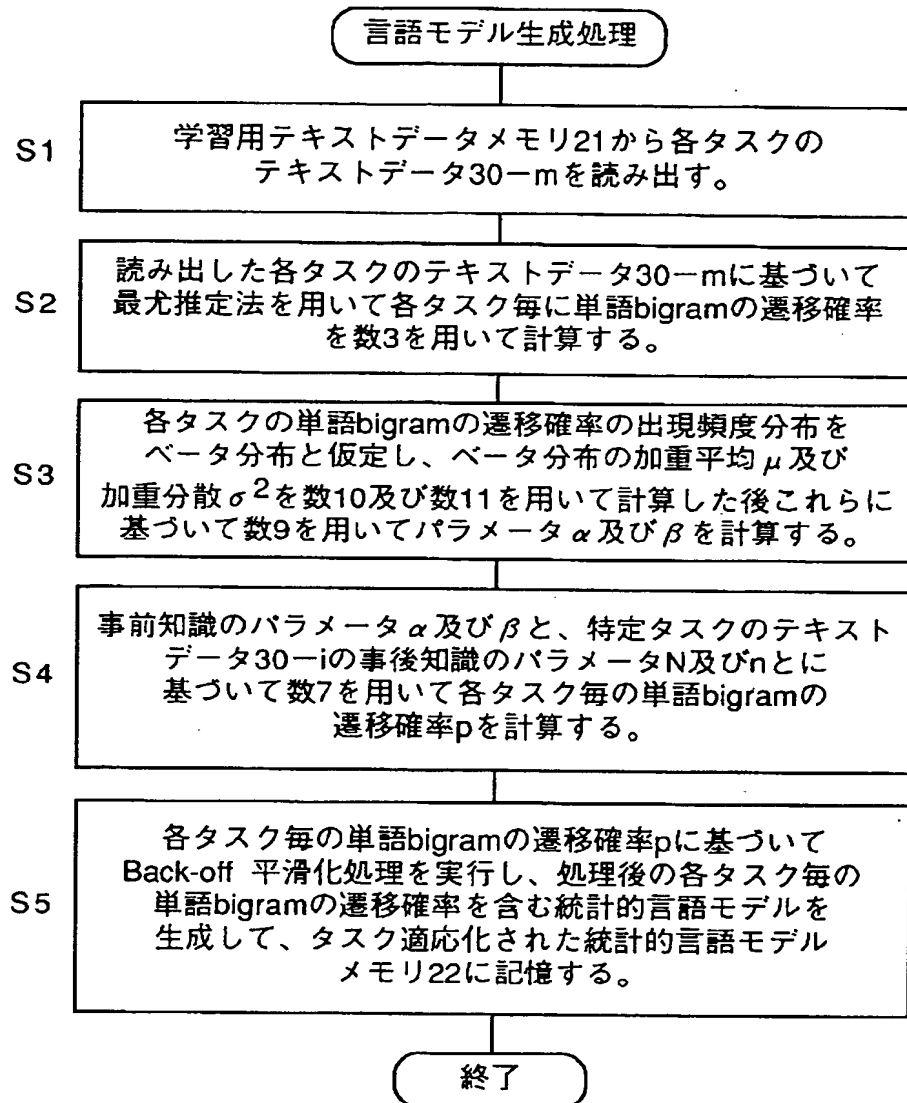
【図3】



【図4】



【図5】



## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-198395

(43)Date of publication of application : 31.07.1998

(51)Int.Cl. G10L 3/00  
G10L 3/00  
G10L 3/00  
G10L 5/06

(21)Application number : 09-002733

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN  
KENKYUSHO:KK

(22)Date of filing : 10.01.1997

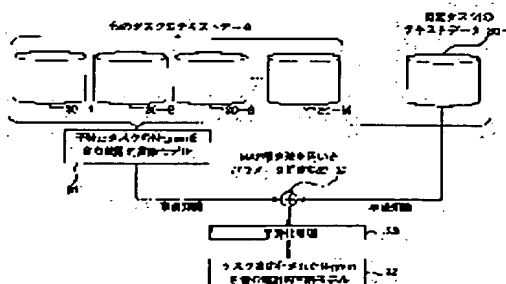
(72)Inventor : MASATAKI HIROKAZU  
KOSAKA YOSHINORI

## (54) STATISTICAL LANGUAGE MODEL GENERATOR AND VOICE RECOGNITION DEVICE

## (57)Abstract:

PROBLEM TO BE SOLVED: To provide a generator which generates a statistical language model which more accurately predicts transition probability and is highly reliable, by providing the device with a means which generates a statistical language model which treats transition probability of N-grm (N is natural number more than one) of word for each task.

SOLUTION: Stable parameter estimation is performed using MAP estimating method in which a prior knowledge is a statistical language model including N-grm which is generated based on text data from 30-1 to 30-M of unspecified tasks, and a postprior knowledge is text data 30-i of an aimed specific task (i), and then smoothing process is performed using Back-off smoothing method. Thus, a statistical language model which includes transition probability of task-adoptive N-grm is provided. Where, text data 30-i of the aimed specific task (i) is one of text data 30-m (m=1, 2,..., M) of unspecified tasks.



## LEGAL STATUS

[Date of request for examination] 10.01.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3027544

[Date of registration] 28.01.2000

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office